



Vessey R.

[Corpus approaches to language ideology.](#)

*Applied Linguistics* (2015)

DOI: 10.1093/applin/amv023

**Copyright:**

This is a pre-copyedited, author-produced PDF of an article accepted for publication in *Applied Linguistics* following peer review. The version of record is available online at:

<http://dx.doi.org/10.1093/applin/amv023>

**Date deposited:**

05/05/2015

**Embargo release date:**

06 June 2017



This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](#)

**Vessey, R. (in press). Corpus approaches to language ideology. *Applied Linguistics*.  
Final accepted pre-proof version.**

## 1. Introduction

Although “language ideology” has been a topic of research for linguists since the 1970s (e.g. Silverstein, 1979), studies have been primarily oriented toward the field of linguistic anthropology. More recently, researchers in other fields have adopted concepts and the literature from the field of language ideologies and have applied new methods. This paper contributes to this burgeoning research area by outlining how corpus linguistics tools and methods can be usefully applied to studies of language ideology.

The field of language ideology was largely defined by the publication of the edited collection *Language Ideologies: Practice and Theory* (Schieffelin, Woolard, and Kroskrity, 1998), which constituted “a first foray into identifying a field of inquiry” (Woolard, 1998: 9). This “field of inquiry” was founded in linguistic anthropology and has been steeped in this domain’s theory and methods. More specifically, ethnographic approaches have tended to predominate, and while such research has produced rich findings, scholars from other disciplines have found that the theory and literature of language ideology have great potential in their own domains. Moreover, some scholars (e.g. Laihonon, 2008; Milani and Johnson, 2008) have drawn parallels between research in their field and that in language ideology, and have posited that the field of language ideology could potentially benefit from closer collaboration with theoretical and methodological developments in their fields. This paper is in keeping with these recent claims: here, it is argued that corpus linguistic methods can be fruitfully applied to the field of language ideology.

This paper proceeds first by outlining the field of language ideology and overviewing some general trends in the research that has been undertaken in the traditional linguistic anthropological domain; then, studies are overviewed that have proposed theoretical and methodological innovations in language ideology research. Next, the paper turns to the important distinction between implicit and explicit manifestations of ideology and the corresponding theoretical and methodological issues. Having identified these issues, the paper then outlines some basic tenets of corpus linguistic theory and method and how these can be usefully applied to studies of language ideology. To illustrate these points, specific

examples of a corpus-assisted discourse study of language ideologies in Canadian newspapers are provided. In the Discussion section, the limitations of the corpus linguistic approach to language ideologies are addressed.

## 2. Language Ideology

In this paper, “language ideology” refers both to a concept and to a field of study. As a concept, the term refers to beliefs about languages (or a particular language) that are shared and that become so well established that their origin is often forgotten by speakers; the beliefs accordingly become naturalised, perceived as common sense, and are socially reproduced. As a field of study, “Language Ideology” (henceforth LI) refers to the body of work that emerged primarily from linguistic anthropologists in the United States, and in particular those associated with the work of Dell Hymes (Milani and Johnson, 2008: 362; Schieffelin, Woolard, and Kroskrity, 1998). The objective in LI research is to understand when and how links are forged between such apparently diverse categories as language, spelling, grammar on the one hand and nation, gender, simplicity, intentionality, authenticity, knowledge, development, power, and tradition on the other (Woolard, 1998: 27). These categories, and the linkages between them, have real effects on the social world; therefore, the study of language ideologies consists of examinations of the broader socio-political contexts in which language ideologies are embedded in order to establish longer term implications for social change (Milani and Johnson, 2008: 373; Wassink and Dyer, 2004: 5).

According to one of the earliest definitions, language ideology refers to “sets of beliefs about language articulated by users as a rationalization or justification of perceived language structure and use” (Silverstein, 1979: 193). This definition and derivations of it are still often used by researchers in the field (e.g. Laihonen, 2008: 669; Stewart, 2012: 190; Wassink and Dyer, 2004). However, some (e.g. Blommaert and Verschueren, 1998) have argued that research should not be singularly focused on “articulated” or explicit manifestations of language ideologies. These researchers have argued that of equal importance are the implicit (“latent”, “immanent”) expressions of these ideologies (see discussion in Woolard, 1998: 9-11). Thus, “[r]epresentations, *whether explicit or implicit*, that construe the intersection of language and human beings in a social world are what we mean by ‘language ideology’” (Woolard, 1998: 3; emphasis added). Language ideologies may be implicit if, for example, they are naturalised and do not require articulation, or they may become explicit in “linguistic

representations” (e.g. Boudreau, 2009), and in particular in “language ideological debates” (Blommaert, 1999). Thus, “[i]deology is variously discovered in linguistic practice itself; in explicit talk about language, that is, metalinguistic or metapragmatic *discourse*; and in the regimentation of language use through more implicit metapragmatics” (Woolard, 1998: 9).

Despite Woolard’s inclusive discussion of the implicit and the explicit nature of language ideologies, she does note that the tension between these different sitings is a recurrent concern to researchers in the field (Woolard, 1998: 6). For example, she notes that Blommaert and Verschueren (1998) posit the importance of naturalised, implicit, “unsaid” ideologies, whereas Briggs (1998) suggests that such an emphasis privileges the analyst’s perspective and may contribute to the analyst’s unintended collusion in reifying the perspective of only a sector of a community (Woolard, 1998: 9). Debates about the “sitings” of language ideology have not been easily dismissed, and researchers (e.g. De Costa, 2010: 220; Griswold, 2010: 407) continue to highlight the distinction between implicitness and explicitness in LI research. Crucially, the distinction has implications not only in terms of theories and definitions (i.e. what language ideology *is*), but also in terms of methodological approach (i.e. how language ideology can be studied). In other words, it is only by establishing whether language ideology occurs in implicit or explicit forms that an appropriate methodology can be established to investigate these forms. If language ideology is understood to occur in both implicit and explicit forms, then accordingly the methodology must enable a researcher to account for both in the data.

To a large degree, the methods that LI researchers have tended to use are oriented towards the theory and methods of linguistic anthropology (see Milani and Johnson, 2008). This is in line with the Hymesian origins of LI (e.g. Hymes, 1974: 31), since theories of language ideologies emerged as a way of enriching and explaining ethnographic data (Woolard, 1998: 14). However, the rich theorisation of language ideology has been increasingly used in disciplines beyond linguistic anthropology. For example, researchers in language and education policy have long been interested in the theories and literature of LI to explain and even to predict the effectiveness of language policy in society (Ricento, 2006: 50). More recently, researchers in conversation analysis (Laihonen, 2008), perceptual dialectology (Stewart, 2012), and phonology (Wassink and Dyer, 2004) have found the explanatory power of language ideologies to be useful in their own work. The theory and literature of LI has been applied to

study subjects as diverse as the language of courtrooms (e.g. Eades, 2012), debates over scripts to represent sign language (Hoffmann-Dilloway, 2011), and the evaluation of language skills in call centres from New Brunswick to Pakistan and in between (e.g. Dubois, Leblanc and Beaudin, 2006; Duchêne, 2009; Rahman, 2009). However, when language ideology is studied in fields where ethnographic data are not in use or appropriate, new methods should be adapted. This is sometimes the case in studies of news media language.

News media are widely seen as an important source of language ideologies, in particular because news discourse is understood to reproduce language and ideologies already in circulation in society (see e.g. Bell, 1991; Johnson and Ensslin, 2007b). Journalists tend to adopt linguistic norms in order to appeal to their “community of coverage” (Cotter, 2010: 25); similarly, the naturalisation of the status quo in newspaper discourse could be argued to lead readers down certain ideological paths (see e.g. Richardson, 2007: 134-5). According to DiGiacomo (1999: 105), the news media have an important function in the reproduction of language ideologies in at least two ways: first, they are places where public figures debate topics directly and indirectly in interviews, articles, and news reports; second, “as literal texts they embody a particular ideology of orthography, syntax, and usage”.

Studies of news media often adopt a discourse approach (see e.g. Cotter, 2001). Discourse analysis has always had a part to play in LI work, not least because of the connections between theories of discourse and ideology in the work of, for example, Michel Foucault (see Woolard, 1998: 7); such approaches have been increasingly used in recent years (e.g. Boudreau, 2009; Milani and Johnson, 2008). In fact, Gal (2006: 388) describes LI as a kind of discourse analysis in which the study of metapragmatic assumptions about the relationship between words, speakers, and worlds provide explanatory power about the effectiveness of verbal action in the society. Milani and Johnson (2008: 365) explain that the traditions of LI and discourse analysis offer “important and potentially *complementary* theoretical and methodological frameworks” (emphasis in original). Indeed, some studies of news media have combined ethnographic and discursive approaches (e.g. Van Hout and Macgilchrist, 2010), which suggests that LI work might fit in easily with discourse approaches to media language. However, discourse approaches to media language have also benefited from an infusion of corpus linguistics theory and methods in the form of “corpus-assisted discourse studies”, or “CADS” (see e.g. Baker, 2006; Partington, 2010; Stubbs, 2001). While CADS

research has already tackled ideology in the media in different forms, there has been little corpus linguistic research on language ideology specifically. What research does exist (e.g. Fitzsimmons-Doolan, 2014; Subtirelu, 2013) has tended to focus only on explicit rather than both explicit and implicit language ideologies. This paper explores the extent to which corpus linguistics can contribute to the study of language ideology in both explicit and implicit forms in news media. The aim is also to explore the advantages and disadvantages of such a combination and to suggest how limitations might be addressed.

In the following sections, this paper outlines some of the primary features of corpus linguistics and how these can contribute to studies of language ideology. In order to illustrate these features, examples are drawn from a larger cross-linguistic corpus-assisted discourse study of language ideologies in Canadian newspapers (Vessey, 2013b). For these examples to be appreciated, it is necessary to provide some basic Canadian background. The following section briefly overviews some socio-political and historical background as well as the data under examination; then, LI findings that were produced through corpus linguistics methods are presented.

### **3. Background and Data**

Canada's official languages are English and French; the status of these languages reflects the fact that the French and English were the first colonisers of the original Canadian territory. However, Aboriginal peoples were the original inhabitants of Canada, and in reality modern day Canadians are far more diverse than a French-English binary would suggest. Indeed, although Canada has never been a country consisting only of English speakers and French speakers, the terms "francophone" (French speaker), "anglophone" (English speaker) and "allophone" (speaker of a language besides English and French, but not an aboriginal language) have long been used as essentialist group labels that enabled the people of Canada to be categorised according to their place in a society that was designed to be French-English bilingual. However, Statistics Canada (the national statistics agency) has recently opted to cease its use of the traditional categories, which apparently no longer reflect the complex linguistic reality of Canada. While the decision by Statistics Canada certainly reflects the broader changes in Canadian society, the replacement of essentialist group labels also indicates a change in frames of reference in the country, which may lead to the gradual devolution of the bilingualism model on which Canada was based in the 1960s and 1970s.

With this changing environment, the present study examines language ideologies in a corpus of 2009 newspaper data.

Two newspapers in English and French were selected, where available, from each of Canada's regions<sup>1</sup>; in addition, national newspapers were selected in English and French<sup>2</sup>. In order to better account for diversity, newspapers were selected from different provinces where possible, or from different cities where a region consisted only of a single province. Despite their ubiquity, no free newspapers (e.g. the *Metro*) were considered, in part because they often tend to be co-ventures with mainstream media partners (Straw, 2010: 89). Notably, nearly all newspapers used for analysis here are the only daily newspaper in the city in which they are produced – a common trend in Canada (Vipond, 2011: 70). The competitive market for daily newspapers tends only to comprise major cities (e.g. Calgary, Winnipeg, Montreal, Vancouver, Toronto, Ottawa), where competition is usually only between broadsheets and tabloids (and often only tabloids are available; e.g. Quebec City) or between the English- and French-reading audiences. Also, all newspapers used in this study are privately-owned and most newspapers belong to sizeable news conglomerates: in 2009 when the data were collected, the *Gazette*, *Calgary Herald*, *Vancouver Sun*, and *National Post* were all owned by CanWest Publishing, one of the largest media stakeholders in Canada, which is also said to support the Conservative Party (Beaty and Sullivan, 2010: 19). Few Canadian newspapers are free of chain ownership (the *Chronicle-Herald* and *Le Devoir* are notable exceptions). Also, no French daily newspapers are produced west of Ontario, where only one daily (*Le Droit*) is produced; only one French daily (*L'Acadie Nouvelle*) is produced in Atlantic Canada.

Data availability and limitations notwithstanding, all articles, editorials, and columns from these newspapers were collected using the news databases *Canadian Newsstand*, *Eureka.cc* and *Actualité Francophone Plus* over the three-week time period 15 June-8 July 2009. The objective in collecting all articles was to account for both explicit and implicit language ideologies. Indeed, this period of time is notable due to the lack of “language ideological debates” (Blommaert, 1999); the “linguistic peace” allowed for a more balanced account of

---

<sup>1</sup> The Canadian Newspaper Association (2009) considers Canada according to five geographic areas: Atlantic Canada (Newfoundland and Labrador, Nova Scotia, New Brunswick, Prince Edward Island), Ontario, the Prairies (Manitoba, Saskatchewan, Alberta), British Columbia and the Yukon, and Quebec.

<sup>2</sup> While there are two “national” English newspapers, no pan-Canadian newspaper exists in French. However, *La Presse* and *Le Devoir* are sometimes considered to be the “national” newspapers in terms of their scope and alignment with Quebec nationalism or a pan-Canadian perspective.

the ideologies that are often inflamed and exaggerated during times of national and linguistic crisis (Cardinal, 2008: 63). Without these more extreme viewpoints, the aim was to ascertain the kinds of explicit and implicit ideologies that circulate in Canadian discourse on a more regular basis.

The complete French corpus consists of a total of 8,759 articles and 3,589,786 words. The English corpus is much larger, consisting of a total of 18,271 articles and 7,524,331 words (see Table 1).

Corpus		Total number of texts	Total running words	Types (distinct words)	% of corpus
Atlantic Canada	L'Acadie Nouvelle	1,421	504,979	32,628	14.07
Quebec	Le Soleil	2,212	778,320	45,684	21.68
Ontario	Le Droit	1,567	600,311	33,842	16.72
Prairies		(no newspapers available)			
BC & Yukon		(no newspapers available)			
National newspapers	La Presse	2,310	1,067,634	55,470	29.74
	Le Devoir	1,249	638,542	45,196	17.79
<b>Total French corpus</b>		<b>8,759</b>	<b>3,589,786</b>	<b>100,286</b>	<b>100%</b>
Atlantic Canada	Moncton Times & Transcript	2,095	956,575	34,704	12.73
	The Halifax Herald	2,453	1,048,651	40,265	13.96
Quebec	The Gazette	1,462	437,310	27,805	5.8
	The Record	188	64,853	9,176	0.86
Ontario	The Toronto Star	1,568	525,760	30,812	7.00
	The Ottawa Citizen	1,825	563,159	29,126	7.49
Prairies	Winnipeg Free Press	1,085	623,717	33,547	8.30
	Calgary Herald	1,476	371,847	24,450	4.95
BC & Yukon	Vancouver Sun	1,205	403,944	24,271	5.38
	Whitehorse Star	501	230,204	17,509	3.06
National newspapers	The Globe and Mail	3,004	1,731,889	56,018	23.05
	The National Post	1,409	493,496	28,061	6.57
<b>Total English corpus</b>		<b>18,271</b>	<b>7,524,331</b>	<b>107,295</b>	<b>100%</b>

Table 1: Data

The French and English corpora are of different sizes but are representative of newspaper readership according to geography across the country. These data were analysed using the corpus suite WordSmith Tools (version 6) (Scott, 2013). Although the original research questions guiding the analysis queried how languages were represented similarly and



differently in English and French, the cross-linguistic data and comparative research design ensured that “representations” were examined according to the salience and content of metalinguistic discussions as well as in the low frequency and absence of such discussions, as explained below (Vessey 2013a, b).

#### 4. Methods and Findings

This section highlights how the frequency, concordance, and “keyword” functions of corpus linguistic programmes can be adapted to multilingual data (in this case, comparable non-parallel corpora) in order to study the language ideologies therein (see also Vessey, 2013a).

The WordList function, which enables researchers to view words according to their rank of frequency or according to alphabetical order, proved to be useful in this study of language ideology because frequent words and phrases may indicate the prominence of certain topics and ways of discussing them. Many words are frequent in a community because they occur in frequent phrases, which are in turn frequent because they are conventional ways of expressing common meaning. However, low frequency items can also help to suggest topics that are taken for granted or avoided in discourses and high and low frequency items can only have meaning when compared against one another (Stubbs, 2001: 221). In other words, high and low frequency (and even absence) can be identified using corpus methods; these have implications for the identification of explicit and implicit language ideologies.

In the English and French corpora of Canadian newspapers, frequency was an important starting point for analysis. One of the first and most obvious findings was the very different frequencies of references to language. Whereas 15.9% of all French newspaper articles over a three-week period contain references to language<sup>3</sup>, over that same three-week period, only 7.86% of all English newspaper articles contain references to language<sup>4</sup>. This suggests that metalanguage (i.e. explicit discussion of language) is considerably more common in French

---

<sup>3</sup> “References to language” comprise the following terms : ANGLAIS, ANGLAISE, ANGLAISES, ANGLICISME, ANGLICISE, ANGLO, ANGLOS, ANGLOPHONE, ANGLOPHONES, BILINGUE, BILINGUES, BILINGUISME, FRANÇAIS, FRANÇAISE, FRANÇAISES, FRANCO, FRANCOIS, FRANCOFONE, FRANCOFONES, FRANCOFONIE, LANGAGE, LANGAGES, LANGAGIER, LANGAGIÈRE, LANGAGIÈRES, LINGUISTIQUE, LINGUISTIQUES, LANGUE, LANGUES

<sup>4</sup> “References to language” comprise the following terms: ANGLO, ANGLOS, ANGLICIZE, ANGLOPHONE, ANGLOPHONES, BILINGUAL, BILINGUALS, BILINGUALISM, ENGLISH, FRANCO, FRANCOFONE, FRANCOFONES, FRANCOFONIE, FRENCH, LANGUAGE, LANGUAGES, LINGUISTIC, LINGUISTICS, MONOLINGUAL, MULTILINGUAL, UNILINGUAL

Canadian newspapers than in English Canadian newspapers. This finding is corroborated by another frequency finding. In the French corpus, references to “French” (*FRANÇAIS/E/S*) are far more frequent than references to “English” (*ANGLAIS/E/S*). In comparison, in the English corpus references to FRENCH occur nearly twice as often as ENGLISH. In both cases, the languages are discussed more in the French corpus than in the English corpus. These frequencies can be compared across languages despite the different sizes of the corpora through the use of normalised frequencies (e.g. frequency per million words; Baker, 2006) (see Table 2)<sup>5</sup>.

Word	Frequency in words per million	Word	Frequency in words per million
FRANÇAIS	301	FRENCH	105
FRANÇAISE	100		
FRANÇAISES	17		
ANGLAIS	126	ENGLISH	56
ANGLAISE	14		
ANGLAISES	1		

**Table 2: Frequencies of *FRANÇAIS/E/S*, *ANGLAIS/E/S*, FRENCH and ENGLISH**

These frequencies appeared to suggest that the English language is naturalised and unmarked in English Canadian newspapers in contrast to comparatively more metalinguistically oriented French Canadian newspapers. However, such conclusions cannot be drawn from frequencies alone; also, such conclusions cannot explain why such discussions of language would be more salient in one community than another. For these reasons, it is crucial to investigate findings further using the qualitative procedures available in corpus linguistics.

The qualitative procedures of corpus linguistics are built to a large degree on the work of John Sinclair (e.g. 1991), who theorised that meaning in language is not created by words used in isolation from one another, but rather from words used in combination. Meaning is often distributed across units larger than individual words, and thus words must be viewed in

<sup>5</sup> Because the larger study explored language and national identity and because the differences between national identity (France vs. French Canadian) and linguistic identity (French speaker vs. French Canadian speaker) are often impossible to distinguish (e.g. both nominal and adjectival forms can be used to refer to language), the cases where ENGLISH/ANGLAIS/E/S and FRENCH/FRANÇAIS/E/S refer to national identity were not excluded. Furthermore, it is impossible to distinguish between these uses of the term using automated methods, even ones as refined as the semantic analysis system of WMatrix (which was attempted). Certainly, some obvious cases could have been excluded, but since such categorisation could not be exhaustive or wholly objective, it was abandoned at early stages.

context (or “co-text”) in order for meaning to be understood (Stubbs, 2001: 100). The Concord tool enables researchers to determine which words collocate with which other words (by default, within five words of the node), thus revealing semantic or discursive relationships. Statistical tests such as Mutual Information (MI) are sometimes used to establish if words co-occur by chance (for an introduction, see e.g. Baker, 2006: 101); MI scores of 3.0 or higher tend to be taken as evidence of statistically significant collocation. Considerations of fixed or semi-fixed phrases, or more generally words that tend to collocate, enable researchers to identify patterns that may suggest linkages between categories. For example, collocation patterns may suggest linkages between linguistic features and social labels – a primary concern for LI researchers.

In order to investigate how metalanguage may pertain to explicit and implicit language ideologies, the data can be approached from a variety of different ways. One example of how the Concord tool served to support this hypothesis was through examining the statistically significant collocates of the terms FRENCH/FRANÇAIS/E/S and ENGLISH/ANGLAIS/E/S (in this case, MI scores are provided, although WordSmith corroborated the significance through other statistical tests).

In the English corpus, ENGLISH collocates with a range of different words. However, the largest category of statistically significant collocates pertains to education. In fact, the collocates SCHOOLS (19 collocations), CLASSES (11 collocations), TEACHER (5 collocations), SCHOOL (25 collocations), STUDENTS (6 collocations), and UNIVERSITY (8 collocations) are all statistically significant (MI scores ranging from 4-7). Similarly, the largest category of statistically significant collocates of FRENCH also pertains to education: IMMERSION (48 collocations), KINDERGARTEN (7 collocations), SCHOOLS (20 collocations), CLASSES (7 collocations), LEARNED (9 collocations), PROGRAMS (6 collocations), SCHOOL (39 collocations), and PROGRAM (9 collocations) (MI scores ranging from 4-9) (see selected concordance lines in Table 3).

(canoe)," she said. Nicole, a French **immersion student** who a  
ldren travel by school bus to French **immersion classes** outsi  
ader March 7, is heading to a French **immersion program** in Jo  
t is under siege over cuts to French **immersion programs**. Two  
vowed to form their own co-op French **immersion kindergarten**  
aduates from Miramichi's lone French- language **school** said g  
tario and **learned** most of his French in **school**, Horwath, 46,  
ls. Lamrock said investing in French **schools** and culture ben  
little or no French." If our French **schools** are prepared to

de Moncton is holding various French <b>classes</b> for all levels
--

**Table 3: Selected concordance lines of FRENCH with education-related collocates**

In the French corpus, there are notable differences. ANGLAIS has a number of statistically significant collocates, but no statistically significant collocates pertain to education. Similarly, although FRANÇAIS/E/S (1601 occurrences) collocates with education-related words (e.g. ÉCOLE (13 collocations), ÉCOLES (13 collocations), ENSEIGNEMENT (7 collocations), ÉTUDES (7 collocations)), these are not statistically significant (see Table 4).

<p>fait nos <b>études</b> primaires en français à Toronto", relate l urs <b>études</b> postsecondaires en français, en Ontario, dans un ioration de l'<b>enseignement</b> du français, en février 2008. Ce l'avenir de l'<b>enseignement</b> en français au Nouveau-Brunswick nçais. Récemment, deux <b>écoles</b> françaises d'Ottawa se sont dé - où, pour sauver les <b>écoles</b> françaises et catholiques, une <b>écoles</b> catholiques de langue française du Centre-Est, ont p 'endroit de l'<b>école</b> de langue française internationale prima fants parlent et entendent du français, c'est à l'<b>école</b>. Il ey Molson fréquentent l'<b>école</b> française. "Je parle français</p>
---

**Table 4: Selected French concordance lines with FRANÇAIS/E/S and education-related collocates**

There are still many more ways in which the data can be examined through corpus tools in order to corroborate findings. The KeyWord tool establishes which words are of statistically significant high frequency (“positive keywords”) or low frequency (“negative keywords”) in comparison with a “comparator” (or “reference”) corpus. The KeyWord tool counts the words in each corpus, measuring their proportion of the overall lexical content of the corpus, and then uses statistical tests (by default, WordSmith uses log likelihood) to determine whether the difference may have occurred by chance (in this case,  $p < 0.000001$ ; however, this can be altered to suit researcher needs). Words that are typical to both the primary and the comparator corpora are eliminated; as a result, only those words whose frequency or scarcity is significant are included on the KeyWord list. Since keywords reveal salient thematic content, these words may have an unusual function in the discourse community from which the data are drawn; thus, they may be of interest in studies of ideology (Kemppanen, 2004: 91).

In many corpus-assisted discourse studies, two corpora are compared against one another in order to derive lists of keywords that distinguish the datasets from one another (see e.g. Baker, 2006: 121-152). However, corpora of different languages cannot be compared against one another directly using the KeyWord tool, and therefore the respective keyword lists must be compared more indirectly (see Vessey, 2013a: 15-20). The KeyWord tool was used here to

establish if articles that contain references to language<sup>6</sup> contain different vocabulary from newspaper articles more generally. If distinctive lexicogrammar is used in newspaper articles that contain references to language, then these items may suggest the contexts in which language issues are seen to be relevant, with potential implications for language ideology. In order to compare language-related and non-language-related articles, the French and English corpora were partitioned: in each, a small subcorpus was created of the articles that contained at least one reference to language. These subcorpora became the primary corpora, which were compared against the whole of each corpus in the same language (i.e. the sum of all articles, henceforth the “comparator corpus”).

One of the most notable trends in the English corpus was the large number of keywords pertaining to education (Table 5). This quantity differed quite dramatically from the number of keywords pertaining to education in the French corpus. These keywords suggest that language education may be more topical in the English Canadian newspapers than the French Canadian newspapers (Table 6).

Positive key word	Frequency	% of words in primary corpus	Comparator corpus frequency	% of words in comparator corpus	Keyness score
SCHOOL	826	0.09	3674	0.05	219.15
STUDENTS	397	0.04	2849	0.02	166.23
STUDENT	149	0.02	1623	0.02	132.50
EDUCATION	311	0.03	1167	0.02	127.17
LITERACY	85		147		105.90
IMMERSION	52		56		93.88
SCHOOLS	225	0.02	848	0.01	91.11
CLASSES	120	0.01	359		75.57
TEACHERS	128	0.01	457		58.21
COURSES	79		234		50.57
LEARNING	125	0.01	993		40.59
TEACHING	77		272		35.79
KINDERGARTEN	54		163		33.51
CAMPUS	54		167		32.17
GRADUATES	61		217		27.95

Table 5: English keywords pertaining to education

Positive Keyword	Frequency	% of words in primary corpus	Comparator corpus frequency	% of words in comparator corpus	Keyness Score
ÉCOLES	105	0.013	347		56.33
ÉLÈVES	145	0.018	563		55.79
ENSEIGNANTS	58		166		39.67

<sup>6</sup> See Notes 3 and 4, above.

L'ÉCOLE	172	0.02	892	0.01	27.71
---------	-----	------	-----	------	-------

**Table 6: French keywords pertaining to education**

However, these keywords can be misleading since a similar proportion of references to education in the primary corpus and the comparator corpus could mean that they cancel each other out and that would explain why there are fewer education-related French keywords. Nevertheless, it is clear that in both the English primary corpus and the English comparator corpus, education-related keywords are more frequent than their equivalents in the French corpus.<sup>7</sup> These findings can be corroborated by downsampling using dispersion plots.

Dispersion plots present the distribution of an item according to its locations in the data, and they can establish consistency and typicality of categories as well as variation and minority trends (Baker, 2010: 39). Thus, dispersion plots can also be used to contextualise high and low frequency items. Dispersion plots can also identify which individual texts contain particularly high or low frequencies of a search word. This can enable a researcher to change from a macro view of the corpus as a whole to a more micro study of an individual text in question. This may be of use to LI researchers who wish to establish greater context for single texts with unusual amounts of metalinguistic commentary. At the same time, LI researchers may wish to identify individual texts that contain little metalinguistic commentary (i.e. few explicit references to languages or language issues) in order to establish if language ideologies are implicit in such cases. For this study, four articles with the most references to language<sup>8</sup> per million words and six articles with the fewest references to language per million words were downsampled from the English and French primary corpora, respectively (see Tables 7 and 8).

<sup>7</sup> My thanks to the anonymous reviewer who highlighted these nuances.

<sup>8</sup> See Notes 3 and 4, above.

- Anonymous. (2009g). Vigilance essential for French. *Toronto Star*, 1 July 2009, p. A17.
- Bélair-Cirino, M. (2009). Le français à Montréal: 90% des francophones sont inquiets. *Le Devoir*, 22 June 2009, p. A1.
- Ferenczy, M. (2009). Broader opportunities. *Ottawa Citizen*, 3 July 2009, p. A9.
- Havrankova, J. (2009). Apprendre le français, un privilège. *Le Devoir*, 22 June 2009, p. A6.
- Howlett, K. (2009). French schools will be available to more students. *The Globe and Mail*, 17 June 2009, p. A12.
- Meurice, P. (2009). Pauvres touristes. *La Presse*, 7 July 2009, p. A13.
- Ravindran, M. (2009). How to speed immigrants' entry into the workforce. *Vancouver Sun*, 30 June 2009, p. A10.
- Rioux, C. (2009). Full bilingue. *Le Devoir*, 3 July 2009, p. A3.

**Table 7: Downsampled articles with the most references to language per million words**

- Anonymous. (2009c). Divine liturgy to be held tomorrow. *Moncton Times & Transcript*, July 4 2009, p. E6.
- Aubry, J. (2009). Le plus grand vin canadien? *Le Devoir*, 19 June 2009, p. B6.
- Blatchford, C. & Leeder, J. (2009). "Did we push her too much?" *The Globe and Mail*, 20 June 2009, p. A1.
- Cornellier, L. (2009). Pierre Falardeau et son Elvis. *Le Devoir*, 27 June 2009, p. E5.
- Lawrence, G. (2009). De l'évanescence à l'effervescence. *Le Devoir*, 20 June 2009, p. D1.
- Le Bouthillier, C. (2009). Le Grand Caraquet – suite. *L'Acadie Nouvelle*, 25 June 2009, p. 13.
- Lussier, M. A. (2009a). De père en flic. *La Presse*, 4 July 2009, p. Cinema 2.
- Mazerolle, B. (2009). The quintessential Canadian. *Moncton Times & Transcript*, 25 June 2009, p. A1.
- Nolen, S. (2009). India's gay community fights for 'dignity'. *The Globe and Mail*, 19 June 2009, p. A16.
- Valpy, M. (2009). The emperor and the tennis pro. *The Globe and Mail*, 27 June 2009, p. A1.
- Vigor, J. C. (2009). Cet art qu'est la composition florale. *Le Devoir*, 20 June 2009, p. D6.
- York, G. (2009). Sterilized, stigmatized. *The Globe and Mail*, 15 July 2009, p. A7.

**Table 8: Downsampled articles with fewest references to language per million words**

The objective of this downsampling was to establish the relationship between the salience of metalinguistic commentary and the implicitness or explicitness of language ideologies.

While space does not permit a full account of the discourse analyses that were undertaken on these downsampled articles (for more details, see Vessey, 2013b), suffice it to say that while two of the four English articles from Table 7 focused on language education issues, none of the four French articles discussed language education at all. In the instances where languages were mentioned in passing (i.e. articles downsampled for their infrequent discussion of language issues), most examples in English tended to reveal instrumental approaches to language, or ideologies of languages as commodities. In these cases, language fluency – and in particular, fluency in English – was represented as an asset to individuals. For example, in one downsampled article the qualities of a Canadian military officer are extolled, including her multilingualism (see Example 1).

**Example 1**

“She got far more high-level attention than a normal RMC [Royal Military College] grad would get,” said a now-retired senior officer who once lobbied for her. But then, he said, she deserved it – she was trilingual (English, French and Portuguese), and she had that marvellous intellect and work ethic.

(Blatchford and Leeder, 2009)

Another article, “India’s gay community fights for ‘dignity’” (Nolen, 2009), represents fluency in English as a positive trait of the educated elite in Indian society. According to this description, being educated and fluent in English allows gay men and lesbians more freedom, both on the Internet and in elite establishments.

In French, the articles with the fewest language references per million words tend to highlight the minority status of the French language in Canada. More specifically, articles mention the English language in passing, seemingly because it is a regular feature in the lives of most French speakers. For example, the occasional necessity for French speakers to be fluent in English is evident in Example 2, where a journalist cites an English-speaking source at length with no French translation, and directs readers’ questions, which must be in English, directly to the source.

**Example 2**

*«We have the perfect climate for wine», disait l’homme, avant de poursuivre:  
«We even have an extra two hours of sunshine over the best terroirs of  
California with an average of 17,4 hours of sunshine!» [...] Pour le reste, il  
faudra poser vos questions, en anglais, directement à Shayn  
(sbjornholm@washingtonwine.org)!*

“We have the perfect climate for wine,” said [Shayn Bjornholm], before continuing: “We even have an extra two hours of sunshine over the best terroirs of California with an average of 17.4 hours of sunshine!” [...] For the rest, send your questions, in English, directly to Shayn (sbjornholm@washingtonwine.org)!

(Aubry, 2009)



In another downsampled example, (Vigor, 2009), the journalist provides the English translation for flowers under discussion. In noting that *plante chenille* is referred to as “Red Hot Cat’s Tail” in English, the journalist suggests that the French name is marginal and that the English name may be more familiar or more useful to French speakers.

In another example, the French-English status quo is represented in a much more negative light. Example 4 is an excerpt from an article in which a Quebec nationalist criticises Quebec filmmakers who make English language films, implying that they are foreign and disloyal to the nation. In this case, the juxtaposition of the English language with an exclusive in-group (“we”, “our”) suggests the extent to which group boundaries are marked by language. Also, the fact that French-speaking filmmakers use English suggests the minoritised status of French when faced with English in Canada (see Example 3).

### Example 3

*«C’est comme si on se fabriquait notre propre culture américaine, pour consommation locale», lance-t-il au sujet de Pascale Picard et des cinéastes québécois qui tournent en anglais.*

“It’s as though we create our own American culture for local consumption”, he hurls out on the subject of Pascale Picard and Quebec filmmakers who shoot in English.

(Cornellier, 2009)

These and other examples suggest that in French Canadian newspapers metalinguistic commentary is linked to the expression of explicit language ideologies because of the necessity for French speakers to be aware of language issues in everyday life. Such awareness is concomitant with community-based beliefs about, evaluations of, and understandings of such language issues (i.e. language ideologies). By contrast, the comparatively small amount of metalinguistic commentary in English Canadian newspapers seems to relate to the implicitness of language ideologies; such ideologies rarely need to be made explicit, given the hegemony of the English language in Canadian society. Although such a hegemony is never “singular” and is continually challenged (Williams, 1973), the role of the English language in Canada is such that it tends to be discussed primarily when

contrasted with other languages or speakers. The focus on language education suggests an interest in languages as social resources that are valuable commodities in English Canadian society. While space does not permit a more comprehensive exploration of these findings, readers are welcome to further investigate these more content-based findings in other publications (Vessey, 2013a, b). For the remainder of this paper, we will turn to a summary of the methods and their potential for use in studies of language ideology.

## **5. Discussion and conclusion**

Results suggested that words of high, low, and statistically significant frequency can help in the identification and exploration of language ideologies within the corpora of news media language. The frequency of references to languages can be used to identify sites where metalanguage is more or less salient. The comparison of metalanguage across corpora – and even corpora of different languages – can also help researchers to explore (potentially ideological) representations of languages within datasets drawn from different social groups. Furthermore, the identification of high and low frequency words helps to establish where language ideologies may prove to be more explicit or implicit. In the corpora examined here, it was revealed that the French Canadian newspapers contain considerably more metalanguage than the English Canadian newspapers; however, the English Canadian newspapers privilege metalanguage about Canada's second official language, whereas the French Canadian newspapers privilege metalanguage about the first official language. Further keyword findings revealed that English Canadian newspapers foreground discussions of language education (e.g. through numerous statistically significant words pertaining to education); in contrast, French Canadian newspapers have much more wide-ranging discussions about language issues. Thus, while frequency findings themselves cannot reveal the “explicitness” or “implicitness” of language ideologies, they do pinpoint potentially fruitful sites for more in-depth investigations.

It was also found that collocation data can provide greater insight into the ways in which languages are represented within sites identified through frequency and statistical significance. Fixed and semi-fixed phrases help researchers to identify patterns in metalanguage (i.e. the ways in which languages are being represented). For example, findings highlighted that ENGLISH and FRENCH collocate with a statistically significant set of words pertaining to education whereas neither ANGLAIS nor FRANÇAIS have statistically significant

collocates pertaining to this subject. Collocation data revealed that, for example, while FRANÇAIS/E/S does collocate with education-related words, these actually comprise a far smaller proportion of the overall collocation trends. Thus, it seemed that FRANÇAIS/E/S was being used in a variety of different ways, suggesting that perhaps in French Canadian newspaper discourse, at least, metalinguistic awareness is embedded in wide-ranging everyday discursive practices. This finding was then corroborated using dispersion plots to downsample whole texts with particularly high and low concentrations of references to language. While English downsampled texts foregrounded practical issues – especially education, but also other issues such as fluency – the French downsampled texts suggested the variety of explicit and implicit ways in which language ideologies seem to figure in everyday issues for French speakers.

The results that emerge from this analysis suggest that these corpus tools can be applied to studies of language ideology in approaches similar to other corpus-assisted discourse studies (e.g. Baker, 2006; Partington, 2010). However, there are limitations to these findings, in particular with relation to concerns about the commensurability of theory and methods and issues pertaining to generalizability.

As mentioned in earlier sections, LI research has primarily tended to draw on ethnographic methods – although this has been changing in recent years. Arguably, a disadvantage of the textual data that tend to be used in corpus linguistics is the lack of contextual richness that is available within large sets of ethnographic data (see e.g. Flowerdew, 2005; Thornbury, 2010). While texts are arguably important sites in which language ideologies can be evidenced, “any form of textual analysis that ties itself too closely to a *structural-linguistic* approach risks losing sight of the broader socio-political contexts (and discourses) from within which those textual meanings are none the less generated” (Milani and Johnson, 2008: 373, emphasis in original). Indeed, “context” itself is an important theoretical concept that is often defined and applied differently in corpus linguistics, ethnography, and discourse analysis. In some discourse research, “context” is taken to mean broad systemic and institutional observations or – problematically, according to Blommaert (2005: 57) – that which is “demonstrably relevant” to texts insofar as those texts show “identifiable traces of social structure”. In contrast, “context” can be taken to mean rather different things in corpus linguistics. For example, Scott and Tribble (2006: 9) identify different layers or “scopes” of context, ranging

from individual words to the context of culture in which a text was written. Other corpus researchers, such as those at Nottingham University (e.g. Adolphs, 2008), have tackled context at the juncture of corpus linguistics and pragmatics and have attempted to account not just for the “wider context of the situation” but also other layers of context pertaining to more local discourse-level negotiation of meaning (Adolphs, 2008: 13). The issue of contextual relevance is also a concern for ethnographers, whose data, despite its advantages, is still decontextualized “at all levels of research endeavours from the data gathering and data configuration, to the analysis of interactional exchanges, or to broader contextualization of societal trends, structures, and values, and, ultimately, to a ‘theory’ of behaviour” (Trueba, 1981: 21). Given these differences, it is perhaps not possible or even desirable to formulate a singular fixed definition of “context” (Goodwin and Duranti, 1992: 2); instead, a shared focus on context and a capitalisation on the advantages of different approaches may allow researchers to access both the etic and emic perspectives that are so crucial within research on language ideology (Trueba, 1981: 32). In other words, perhaps it is a combination of corpus linguistics and ethnography that might enhance research on language ideology. Several researchers have already combined corpus and ethnographic methods (e.g. Gardner, 2008; Koester, 2010), but the combination has not yet been applied to language ideology research more specifically.

Since the research outlined here did not undertake any ethnographic research alongside the corpus and discourse work, there are certainly contextual layers absent from this analysis. Furthermore, the three-week time period and lack of diachronic data means that there is little scope for assessing change over time. With a diachronic corpus, frequency, collocation and concordance analysis could help to establish if the terms and phrases discussed here continue to be salient and used in similar ways over longer time periods. Such findings could be useful, for example, to chart the changing (and changed) meanings of social categories such as “francophone”, “anglophone” and “allophone”, which continue to be used only insofar as these sociolinguistic categories are understood to remain meaningful – something that seems to be changing in recent years, as discussed above. Finally, although it was found that corpus tools can contribute to studies of language ideology in news media, it remains to be determined if such an approach could be usefully applied to other data types (however, see Fitzsimmons-Doolan, 2014). Nonetheless, this study has demonstrated some of the advantages of applying corpus approaches to sets of multilingual data (in this case,

comparable corpora in English and French), suggesting possibilities for future research on language ideology.

In conclusion, although this paper has focused primarily on how corpus linguistics can contribute to the study of implicit and explicit language ideologies, it should also be stressed that the rich literature of language ideology has much to contribute to the field of corpus linguistics. A closer alignment of socially-oriented corpus linguistic research with LI would surely serve to enhance findings, as some researchers (e.g. Fitzsimmons-Doolan, 2014; Subtirelu, 2013) have begun to demonstrate.

## References

- Adolphs, Svenja (2008). *Corpus and context: Investigating pragmatic functions in spoken discourse*. Amsterdam: John Benjamins.
- Baker, Paul (2006). *Using corpora in discourse analysis*. London & New York: Continuum.
- Baker, Paul (2010). *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.
- Beaty, Bart and Sullivan, Rebecca (2010). Introduction: Contexts of popular culture. In B. Beaty, D. Briton, G. Filax and R. Sullivan (eds), *How Canadians communicate III: Contexts of Canadian popular culture* (pp. 11-34). Edmonton: Athabasca University Press.
- Bell, Allan (1991). *The language of news media*. London: Blackwell. Chapter 3: The production of news language
- Blatchford, Christie and Leeder, Jessica (2009). "Did we push her too much?" *The Globe and Mail*, 20 June 2009, p. A1.
- Blommaert, Jan (ed) (1999). *Language ideological debates*. Berlin: Mouton de Gruyter.
- Blommaert, Jan (2005). *Discourse: A Critical Introduction*. Cambridge: Cambridge University Press.
- Blommaert, Jan and Verschueren, Jef (1998). The role of language in European nationalist ideologies. In Bambi B. Schieffelin, Kathryn A. Woolard, and Paul V. Kroskrity (eds), *Language Ideologies: Practice and theory*, 189-210. Oxford: Oxford University Press.
- Boudreau, Annette (2009). La construction des représentations linguistiques: le cas de l'Acadie. *The Canadian Journal of Linguistics/ La revue canadienne de linguistique*, 54 (3), 439-459.
- Briggs, Charles L. (1998). "You're a liar – you're just like a woman!": Constructing dominant ideologies of language in Warao Men's gossip. In Bambi B. Schieffelin, Kathryn A. Woolard, and Paul V. Kroskrity (eds), *Language Ideologies: Practice and theory*, 229-255. Oxford: Oxford University Press.
- Cardinal, Linda (2008). Linguistic peace: A time to take stock. *Inroads*, 23, 62-70.

- Cotter, Colleen (2001). Discourse and media. In Deborah Schiffrin, Deborah Tannen and Heidi E. Hamilton (eds), *The handbook of discourse analysis*, 416-436. Oxford: Blackwell Publishers.
- Cotter, Colleen (2010). *News talk: Investigating the language of journalism*. Cambridge: Cambridge University Press.
- De Costa, Peter I. (2010). Language ideologies and standard English policy in Singapore: responses of a 'designer immigrant' student. *Language Policy*, 9 (3), 217-239.
- DiGiacomo, Susan M. (1999). Language ideological debates in an Olympic city: Barcelona 1992-1996. In Jan Blommaert (Ed.), *Language ideological debates*, 105-142. Berlin: Mouton de Gruyter.
- Dubois Lise, LeBlanc, Mélanie, and Beaudin, Maurice (2006). La langue comme ressource productive et les rapports de pouvoir entre communautés linguistiques. *Langage & société*, 118 (4), 17-41.
- Duchêne, Alexandre (2009). Marketing, Management and Performance: Multilingualism as Commodity in a Tourism Call Centre. *Language Policy*, 8 (1): 27-50
- Eades, Diana (2012). The social consequences of language ideologies in courtroom cross-examination. *Language in Society*, 41 (4), 471-497.
- Eckert, Penelope (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12 (4), 453-476.
- Fitzsimmons-Doolan, Shannon (2014). Using lexical variables to identify language ideologies in a policy corpus. *Corpora*, 9 (1), 57-82.
- Flowerdew, Lynne (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies. *English for Specific Purposes*, 24, 321-332.
- Gal, Susan (2006). Language, its stakes and its effects. In Robert E. Goodin and Charles Tilly (eds), *The Oxford Handbook of Contextual Political Analysis*, 376-391. Oxford: Oxford University Press.
- Gardner, Sheena (2008). Integrating ethnographic, multidimensional, corpus linguistic and systemic functional approaches to genre description. In E. Steiner and S. Neumann (eds), *Data and Interpretation in linguistic analysis*. 19th European SFL Conference

- (July 2007), Universität des Saarlandes. Available: <http://scidok.sulb.uni-saarland.de/sulb/portal/esflcw/>.
- Goodwin, Charles & Duranti, Alessandro (1992). Rethinking context: An introduction. In A. Duranti & C. Goodwin (eds), *Rethinking context* (pp. 1-42). New York: Cambridge University Press.
- Griswold, Olga V. (2011). The English you need to know: Language ideology in a citizenship classroom. *Linguistics and Education*, 22 (4), 406-418.
- Hoffmann-Dilloway, Erika (2011). Writing the smile: Language ideologies in, and through, sign language scripts. *Language & Communication*, 31 (4), 345-355.
- Hymes, Dell (1974). *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia: University of Pennsylvania Press.
- Johnson, Sally and Ensslin, Astrid (eds). (2007). *Language in the media: Representations, identities, ideologies*. London: Continuum.
- Kemppanen, Hannu (2004). Keywords and ideology in translated history texts: A corpus-based analysis. *Across Languages and Cultures*, 5 (1), 89-106.
- Koester, Almut (2010). Building small specialised corpora. In A. O’Keeffe and M. McCarthy (eds), *The Routledge Handbook of Corpus Linguistics* (pp. 66-79). New York: Routledge.
- Laihonen, Petteri (2008). Language ideologies in interviews: A conversation analysis approach. *Journal of Sociolinguistics*, 12 (5), 668-693.
- Milani, Tommaso M. and Johnson, Sally (2008). CDA and language ideology: Towards a reflexive approach to discourse data. In Ingo H. Warnke and Juergen Spitzmuller (eds), *Methoden der Diskurslinguistik Sprachwissenschaftliche Zugänge zur transtextuellen Ebene*, 361-384. Berlin: Mouton de Gruyter.
- Partington, Allan (2010). Modern diachronic corpus-assisted discourse studies (CD-CADS) on UK newspapers: an overview of the project. *Corpora* 5 (2), pp. 83–108.
- Rahman, Tariq (2009). Language ideology, identity and the commodification of language in the call centres of Pakistan. *Language in Society*, 38 (2), 233-258.
- Ricento, Thomas (2006). Americanization, language ideologies, and the construction of European identities. In Clare Mar-Molinero and Patrick Stevenson (eds), *Language*



- ideologies, policies, and practices: Language and the future of Europe*, 44-57. Houndmills, Basingstoke: Palgrave Macmillan.
- Richardson, John E. (2006). *Analysing newspapers: An approach from critical discourse analysis*. New York: Palgrave MacMillan.
- Schieffelin, Bambi B., Woolard, Kathryn A. and Kroskrity, Paul V. (eds), *Language Ideologies: Practice and theory*. Oxford: Oxford University Press.
- Scott, Mike and Tribble, Christopher (2006). *Textual Patterns. Key words and corpus analysis in language education*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Scott, Mike (2013). *WordSmith Tools Version 6*. Liverpool: Lexical Analysis Software.
- Silverstein, Michael (1976). Language Structure and Linguistic Ideology. In *The Elements: A Parasession on Linguistic Units and Levels*, ed. Paul R. Clyne, William F. Hanks, and Carol L. Hofbauer, pp. 193-247. Chicago: Chicago Linguistic Society.
- Sinclair, John (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Stewart, Christopher Michael (2012). Mapping language ideologies in multi-ethnic urban Europe: the case of Parisian French. *Journal of Multilingual & Multicultural Development*, 33 (2), 187-202.
- Straw, Will (2010). Hawkers and public space: Free commuter newspapers in Canada. In B. Beaty, D. Briton, G. Filax and R. Sullivan (eds), *How Canadians Communicate III* (pp. 79-94). Edmonton: Athabasca University Press.
- Stubbs, Michael (2001) *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.
- Subtirelu, Nicolas Close (2013). 'English... it's part of our blood': Ideologies of language and nation in United States Congressional discourse. *Journal of Sociolinguistics*, 17 (1), 37-65.
- Thornbury, Scott (2010). What can a corpus tell us about discourse? In A. O'Keeffe and M. McCarthy (eds), *The Routledge Handbook of Corpus Linguistics* (pp. 270-287). New York: Routledge.
- Trueba, Henry T. (1981). The meaning and use of context in ethnographic research: Implications for validity. *NABE Journal*, VI (2&3), 21-34.

- Van Hout, Tom and Macgilchrist, Felicitas (2010). Framing the news: an ethnographic view of business newswriting. *Text & Talk*, 30 (2), 168-191.
- Vessey, Rachelle (2013a). Challenges in cross-linguistic corpus-assisted discourse studies. *Corpora*, 8 (1), 1-26.
- Vessey, Rachelle (2013b). *Language ideologies and discourses of national identity in Canadian newspapers: A cross-linguistic corpus-assisted discourse study*. Unpublished PhD thesis. Queen Mary, University of London.
- Vipond, M. (2011). *The mass media in Canada*. Toronto: James Lorimer & Company Ltd.
- Wassink, Alicia Bedford and Dyer, Judy (2004). Language ideology and the transmission of phonological change changing indexicality in two situations of language contact. *Journal of English Linguistics*, 32 (3), 3-30.
- Williams, R. (1973). Base and superstructure in Marxist cultural theory. *New Left Review*, 82, 1-14.
- Woolard, Kathryn A. (1998). Introduction: Language ideology as a field of inquiry. In Bambi B. Schieffelin, Kathryn A. Woolard, and Paul V. Kroskrity (eds), *Language Ideologies: Practice and theory*, 3-50. Oxford: Oxford University Press.